MODIS Snow Cover Mapping Decision Tree Technique: Snow and Cloud Discrimination

GEORGE RIGGS[1], DOROTHY K. HALL[2]

ABSTRACT:

Accurate mapping of snow cover continues to challenge cryospheric scientists and modelers. The Moderate-Resolution Imaging Spectroradiometer (MODIS) snow data products have been used since 2000 by many investigators to map and monitor snow cover extent for various applications. Users have reported on the utility of the products and also on problems encountered. Three problems or hindrances in the use of the MODIS snow data products that have been reported in the literature are: cloud obscuration, snow/cloud confusion, and snow omission errors in thin or sparse snow cover conditions. Implementation of the MODIS snow algorithm in a decision tree technique using surface reflectance input to mitigate those problems is being investigated. The objective of this work is to use a decision tree structure for the snow algorithm. This should alleviate snow/cloud confusion and omission errors and provide a snow map with classes that convey information on how snow was detected, e.g. snow under clear sky, snow under cloud, to enable users' flexibility in interpreting and deriving a snow map. Results of a snow cover decision tree algorithm are compared to the standard MODIS snow map and found to exhibit improved ability to alleviate snow/cloud confusion in some situations allowing up to about 5% increase in mapped snow cover extent, thus accuracy, in some scenes.

Keywords: MODIS, snow cover, decision tree

INTRODUCTION

MODIS daily snow cover maps have been used in many applications, e.g. compiling snow depletion curves, input to hydrologic models, studying change in snow cover, etc., with reasonably successful results. Though the MODIS snow maps can be accurate, up to about 93% (or higher) accuracy (Hall and Riggs, 2007), users have reported errors of commission and omission, and limitations imposed by cloud cover including snow/cloud confusion. A common impediment to snow cover mapping with visible sensors is cloud cover because it prevents viewing the surface and clouds can be confused with snow resulting in decreased accuracy of the snow map. Though cloud cover cannot be eliminated, we can attempt to improve the ability to discriminate between snow and cloud and possibly detect snow underlying a thin cloud cover. We have developed a MODIS snow cover algorithm using a decision tree technique that improves the ability to detect snow cover under thin cloud cover and reduces snow/cloud confusion allowing for a more accurate mapping of snow cover extent. The snow cover map from the decision tree technique also contains more classes of features thus increases information content relative to what is available in the standard MODIS snow cover products.

[1]Science Systems and Applications, Inc. Lanham, MD 20706
[2]Cryospheric Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, MD 20771

DECISION TREE TECHNIQUE

The decision tree technique is a very useful and flexible technique for classification of satellite observations (Friedl and Brodley, 1997) and has been used successfully to classify observations in a variety of studies (e.g. Pouliot et al., 2009). A decision tree algorithm has the ability to classify data using continuous or discrete variables, is free of statistical assumptions (Pouliot et al., 2009) and can reveal hierarchical and nonlinear relationships in the variables (Hansen et al., 1996). A decision tree is constructed of rules successively applied to the input data to split the data into two groups. Each split, referred to as a node, results in a more homogenous group of data (Simard et al., 2000.) Rules apply different tests or screening of data along a branch and may use the same or similar tests in other branches and can be developed for any number of classes or thematic maps as output to achieve very accurate classifications. The objective here is to develop a basic decision tree to improve snow/cloud discrimination based on explicit characteristics discerned in the data. A rigorous development of a decision tree by growing and pruning and statistical analysis of purity at nodes was not undertaken. The output of the decision tree is a snow cover map that decreases snow/cloud confusion errors and contains more information compared to the standard Collection 5 MODIS snow products.

MODIS SNOW COVER DECISION TREE

The MODIS snow cover decision tree (MOD_DT) is shown in Fig. 1. The MOD_DT emulates the standard MODIS snow algorithm (MOD10), the light blue nodes and classes in Fig. 1, and adds nodes and classes, blue nodes and classes in Fig. 1, to improve discrimination between snow and cloud in some situations and outputs a unique class at end of each branch. The result of each node in the decision tree is a unique data value. Confusion of snow and clouds, specifically when the cloud mask erroneously flags snow as cloud, decreases the accuracy of the snow map. The version of the MOD_DT presented here is structured to improve snow mapping by including nodes (tests) that alleviate snow/cloud confusion in some situations based on characteristics discerned from analysis of the data. Development and testing of the MOD_DT has been done using the MODIS daily surface reflectance product (MOD09GA) as input, which makes the input different from the at-satellite reflectance input (MOD02HKM) used in the standard MODIS snow algorithm (MOD10) (Riggs et al., 2006). The surface reflectance product is atmospherically corrected and contains cloud flags from the cloud mask product MOD35, and an internal MOD09 cloud flag, but only the MOD35 cloud flag is used in MOD10 and in this version of MOD_DT. To avoid confounding results in comparing MOD10 and MOD10_DT snow maps, the MOD10 and MOD_DT algorithms were run using the same MOD09GA input so that the output snow maps contain only differences and similarities attributable to the MOD_DT algorithm. Comparative analysis of the two snow maps combined with visual interpretation of the imagery was used to evaluate the results of the MOD_DT and to develop the nodes.

In the MOD_DT algorithm (Fig. 1) nodes and results common in both MOD10 and MOD_DT are shown in light blue; nodes and results unique to MOD_DT are shown in blue. The MOD_DT algorithm is structured as two branches, one for clear sky condition, and the other for cloudy sky conditions. The clear sky branch is common to both MOD10 and MOD_DT however MOD_DT has unique values for each node result. The MOD35 cloud mask reports four cloud conditions, confident clear, probably clear, uncertain clear and cloudy (Ackerman et al., 2010). In both algorithms a cloud mask flag of clear or probably clear is interpreted as clear, and cloud is interpreted as cloud. For clear sky observations the two algorithms are essentially the same, however MOD_DT has a different class output at terminal nodes. Branches for cloud observation are different in each algorithm. MOD10 applies the cloud mask and if cloud mask is set to cloud the algorithm terminates. In MOD_DT processing continues for a cloud observation through several nodes to determine if the observation is a snow/cloud confusion situation that could possibly be resolved. The purpose of the nodes in the cloud branch is to resolve snow/cloud confusion to improve quality of the snow map. The cloud branch of MOD_DT has nodes for

snow detection (CDL1 and CDL2) and restoral nodes (CDL3 and CDL4) that can change a snow result back to cloud. The purpose of the restoral nodes is to avoid snow commission errors of classifying cloud as snow. Nodes are described in the following section.

The output of the MOD_DT is a classified snow map reached through branches of the algorithm. The snow map is similar to the MOD10 map but includes information conveyed in the unique classification value that corresponds to the branch result. The output of specific instances of snow or other feature has intrinsic quality assessment (QA) information in that the result was from good view situation, e.g. clear view high reflectance or that snow is observed but that reflectance was low, or snow from a snow/cloud confusion situation which may suggest increased uncertainty in result (see Riggs et al., 2006 for further information on QA).
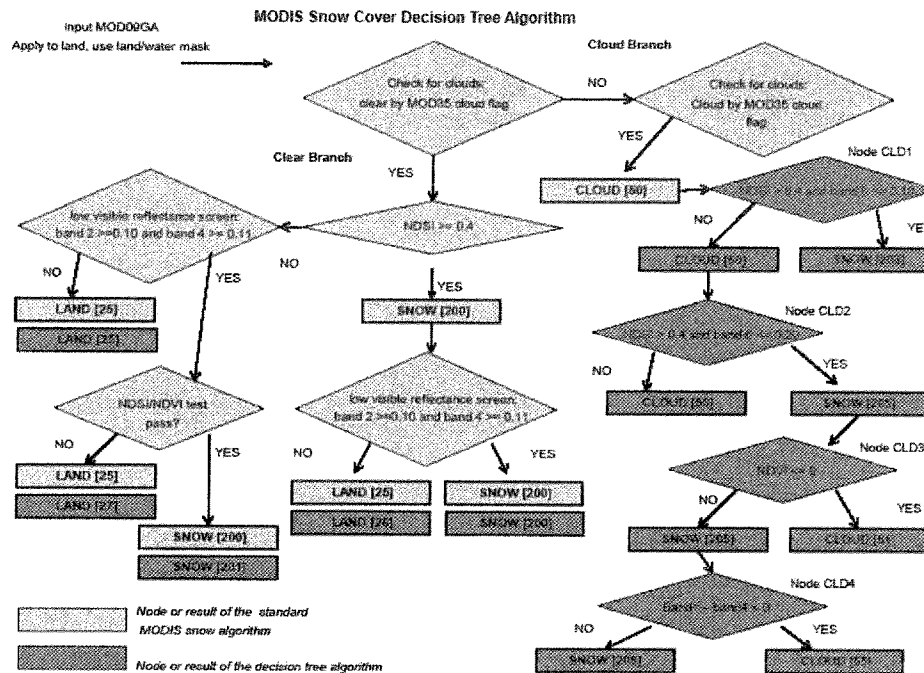

Figure 1. Flow diagram of the MOD_DT algorithm. Nodes and classes in light blue are the same as in MOD10 and those in blue are unique in the MOD_DT algorithm.

DISCUSSION AND RESULTS

Comparative analysis of the MOD10 and MOD_DT along with visual interpretation of the MOD09GA imagery was done for the first ten days of 2003 for tile h11v04, in the MODIS sinusoidal projection covering a region west of the Great Lakes (Fig. 2). Over that time period the extent of cloud cover over the tile varied from about 60-90% and observable snow covered ranged from about 5-20% (Table 1). Situations of cloud cover type, extent and snow cover extent were good for developing and testing the MOD_DT especially the snow/cloud discrimination tests (CDL3 and CDL4 Fig 1.). The series of MOD09GA false color images (MODIS bands 1, 4, 6 as RGB image), MOD10 and MOD_DT snow maps are shown in Figure 3. There were usually 2-4 MODIS swaths mapped into the tile each day, except for on 6 January when only one swath was mapped thus the missing data that day (Fig. 3). A scan line was missing on 4 January (Fig. 3). There was an unusually large amount of thin cloud or aerosol thickness on 8 January which causes the yellow hued image of MOD09GA (Fig. 3) however, that aerosol thickness had a minor impact

on the snow algorithms as they output snow maps that are reasonably good (Fig. 3). Overall, snow cover extent mapped in both the MOD10 and MOD_DT snow maps is very similar on all days in the time series (Figure 3). Snow maps are nearly identical under cloud free conditions however they differ in some cloud cover conditions because of the nodes in MOD_DT that are applied to alleviate snow/cloud discrimination in some situations. The MOD_DT snow cover map is improved over the standard MOD10 snow map because the information content of the snow map is increased with the result of each branch of the decision tree having a unique value, and decreased snow/cloud confusion, allows for more accurate mapping of snow cover extent. Nodes in the clear sky branch of MOD_DT are the same as applied in MOD10 except that in MOD_DT there are three classes of snow free land, classes 25, 26, 27, depending on the terminal
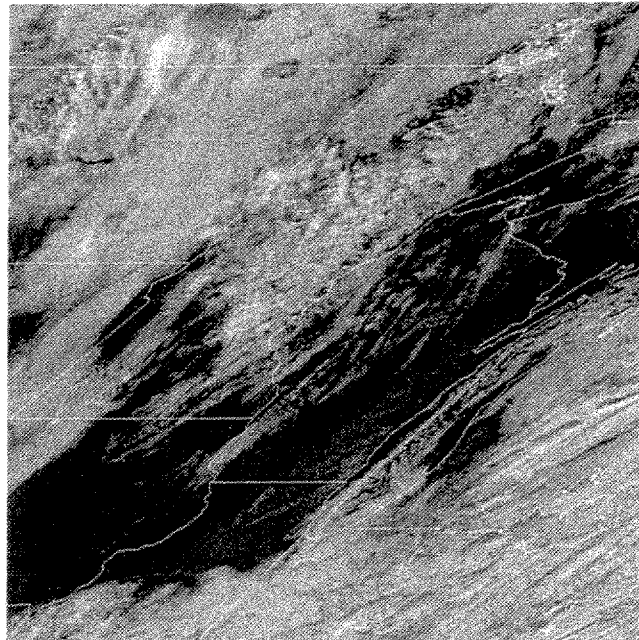

Figure 2. Color image of MOD09GA 2003003 tile h11v04, MODIS bands 1, 4, 6 (RGB). State lines are mapped to provide a visual geolocation reference, in this figure and for Figure 3 images. This band (color) combination is used to make snow cover appear in yellow, land in very dark blue hues, and clouds in white except that   some ice type of clouds which are similar to snow have yellow hued color.

node compared to the single land class, 25, in MOD10. In MOD10 snow is single class but in MOD_DT there are two snow classes, 200 and 201 with the latter indicating that snow was determined by the NDSI and NDVI test for snow mixed with vegetation. That normalized difference snow index (NDSI) and normalized difference vegetation index (NDVI) test is the same in MOD10 and MOD_DT. The node of low visible reflectance in the clear branch (Fig. 1) functions to avoid snow commission errors in low solar illumination conditions.

Observations flagged as cloud are processed differently in MOD10 and MOD_DT. In MOD10 processing ends with the observation classed as cloud. In the cloud branch of MOD_DT (Fig. 1) a cloud observation is processed through several nodes to reduce snow/cloud confusion in situations where the cloud mask erroneously flagged snow cover as cloud. If an observation is flagged as cloud the observation is tested for characteristics indicative of snow in the MOD_DT nodes CDL1, CDL2, CDL3 and CDL4 (Fig. 1) in the cloud branch of the tree.

The next node is a test on the NDSI; if a cloud observation has a high NDSI value and a low near infrared reflectance then it is classed as snow, assigned to class 203 in CDL1 (Fig. 1) and appears in orange color in Fig. 3 images. The objective of this test is to detect snow under thin translucent cloud; this class result is given in column 4 of Table 1. If the cloud observation is

unchanged it goes to node CDL2 (Fig. 1) and tested again for a snow signal of high NDSI value and a more liberal threshold of 20% or less for near infrared reflectance. The objective of this test is to discriminate in situations of an erroneous cloud flagging of snow as cloud. If this test is passed an observation is classed as snow, snow class of 205 (Fig.1) and appears as pale blue in the MOD_DT snow maps in Figure 3. If this test is passed as snow then the observation is tested to affirm the snow result. Development and analysis of the MOD_DT at this node found that clouds were frequently misclassified as snow. Inspection of plots of reflectance curves, and NDSI and NDVI values of these and other classes in scenes led to the application of the CDL3 test (Fig. 1) to avoid misclassifying cloud as snow. A pure cloud observation would have a negative NDVI so if an observation has a negative NDVI it is classified as cloud, or class 51; cells of this class appear in yellow in the MOD_DT snow maps (Fig. 3). A snow result is then again passed to node CDL4 to affirm the snow result. Inspection and comparison of reflectance plots of observations that failed the CDL3 test with observations in a scene that were certainly cloud suggested that observations that were still flagged as cloud by cloud mask and that visually appeared to be cloud, had a band 1 – band 4 difference of < 0.0 which was applied as node CDL4. If an observation passed this test it was classed as cloud, class 55 and appears as light pink in Fig. 3. If an observation failed it was classified as snow, or class 205, and appears as pale blue in the MOD_DT snow maps in Fig. 3.

The area of decreased snow/cloud confusion, cloud error alleviated resulting in increased detection of snow cover, from the MOD_DT is usually in range of 0-5% of land area in a tile (Table 1), though a small extent relative to the entire scene, it does improve the snow map. From a snow/cloud confusion perspective there was a 0 – 8% decrease in cloud cover, based on cloud cover extent in a scene (Table 1, column 5).

Table 1. Summary of snow and cloud cover, and reduction of snow/cloud confusion contributing to increased amount of snow cover detection over the period 1-10 January 2003. Percentage of snow (under clear sky) and cloud are based on all land area in the tile. Percentage of snow/cloud changed based on total amount of cloud.

Date	Cloud cover (%)	Snow cover (%)	Snow detected under translucent cloud; snow class [203] (%)	Snow discriminated from cloud; snow class [205] (%)	Total snow cover from MOD_DT (%) and increase in snow cover
2003001	87.83	1.73	0.10	0.78	2.5 (+ 0.77)
2003002	80.73	5.56	0.13	2.85	7.99 (+ 2.43)
2003003	60.45	18.8	0.35	3.57	21.17 (+ 2.37)
2003004	85.80	6.53	0.23	1.00	7.58 (+ 1.05)
2003005	97.1	2.65	0.04	0.17	2.86 (+0.21)
2003006	78.0	8.02	0.36	3.17	10.78 (+2.76)
2003007	65.58	15.65	1.62	4.90	19.92 (+4.27)
2003008	48.00	20.85	1.94	3.45	23.44 (+2.59)
2003009	71.21	6.08	0.23	1.26	7.14 (+1.06)
2003010	65.43	9.07	0.33	7.55	14.23 (+5.16)

The cloud branch path is responsible for alleviating snow/cloud confusion in the series being analyzed. There was up to about 5% increase in total snow cover in some scenes (Table 1). Results varied with the snow and cloud conditions in each scene. Detail of a situation with results common along this path is shown in Figure 4 for 3 January 2003 which is the lower left corner of the tile in Figure 3. Visual interpretation of that area reveals that some of the snow cover was correctly classified as snow from cloud through the CDL2, CDL3 and CDL4 nodes (pale blue) and that some of the observations classed as cloud from those nodes, yellow and pale pink colors, are

visually interpreted as snow in MOD09GA image yet were classed as cloud. That results from clouds in other areas of the scene and in other scenes that were observed to have similar characteristics. We experimented with variations to the nodes and path to balance the ability to decrease snow/cloud confusion in situations where visual interpretation revealed erroneous cloud masking of snow cover without causing concomitant errors of snow commission on clouds in other situations in a scene. Reflectance plots (not shown here) were made of the classes in a scene at each node of the decision tree to aid evaluation of our ability to improve snow mapping and determine if and where concomitant errors occurred.

Flexibility of the decision tree technique for adding nodes and splitting classes was exercised by adding nodes based on examination of the reflectance plots and other characteristics of a surface feature (class). Comparative analysis of reflectance plots of classes in a scene was done to search for characteristics that could be used to further separate observations in a class into other classes. If a reflectance feature or characteristic, e.g. ratios or band differences, of a surface feature could be found to enable separation of a feature from other features then a node could be defined to do that. The focus here has been on resolving snow/cloud confusion in situations where the cloud mask erroneously flags snow cover as cloud in situation shown in Figure 4. Earlier in development of the MOD10 snow algorithm there was an attempt to improve snow/cloud discrimination through use of the MOD35 level 2 cloud mask and individual cloud detection tests from the MOD35 product (Riggs et al., 2002). That attempt alleviated snow/cloud confusion at edges of snow covered areas, similar to that shown in Figure 4, but concomitantly missed large areas of cloud and erroneously classified them as land. Experience with development of the MOD_DT suggests courses of future research into improved snow cover mapping. Possibilities include: bringing in the MODIS land cover type as a factor in snow mapping, using the MOD09GA internal cloud flag in combination with the MOD35 cloud flag to improve cloud masking relative to snow/cloud discrimination. Some preliminary research has been done using the combined cloud masks; cloud masking was improved by using both masks to map cloud in a scene. Another option is to investigate reducing snow/cloud confusion at the MOD10 level 2 processing with a decision tree using the MOD35 individual cloud mask test flags and reflectance data to improve snow/cloud discrimination.
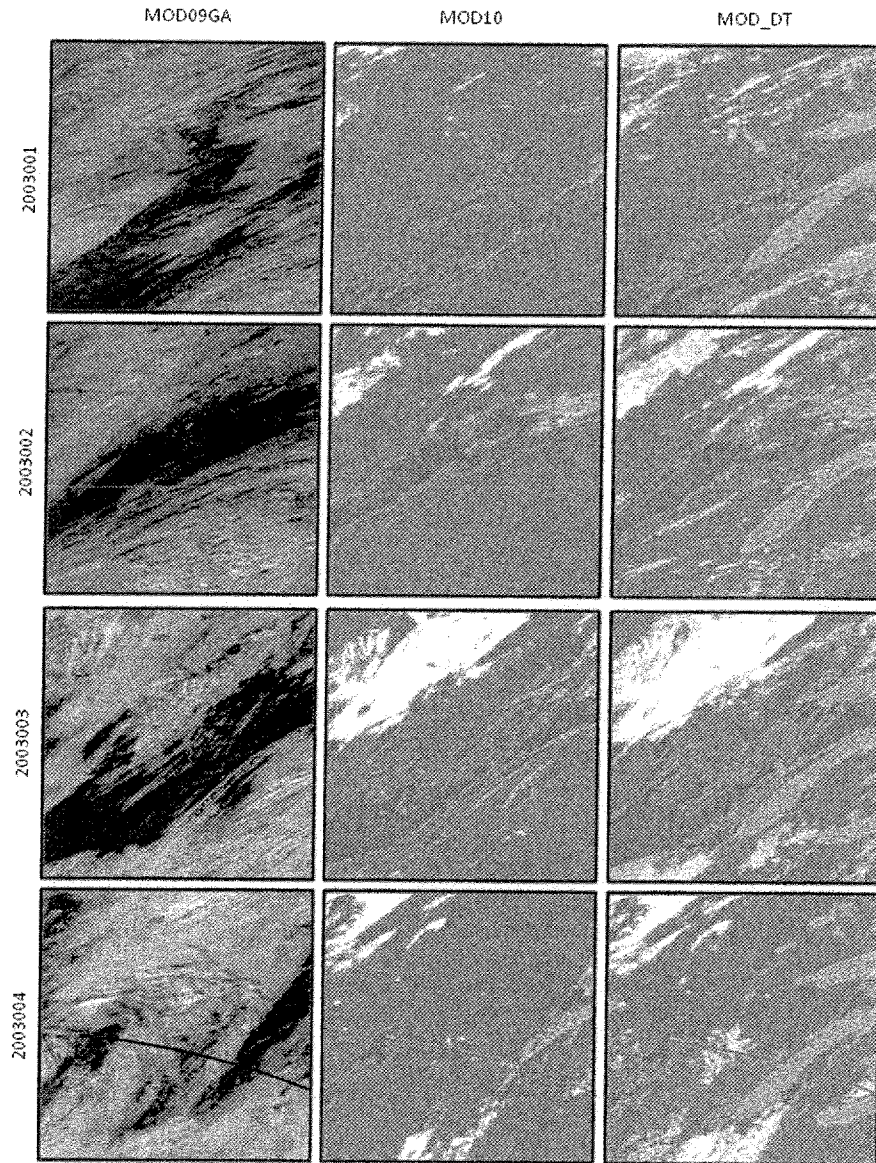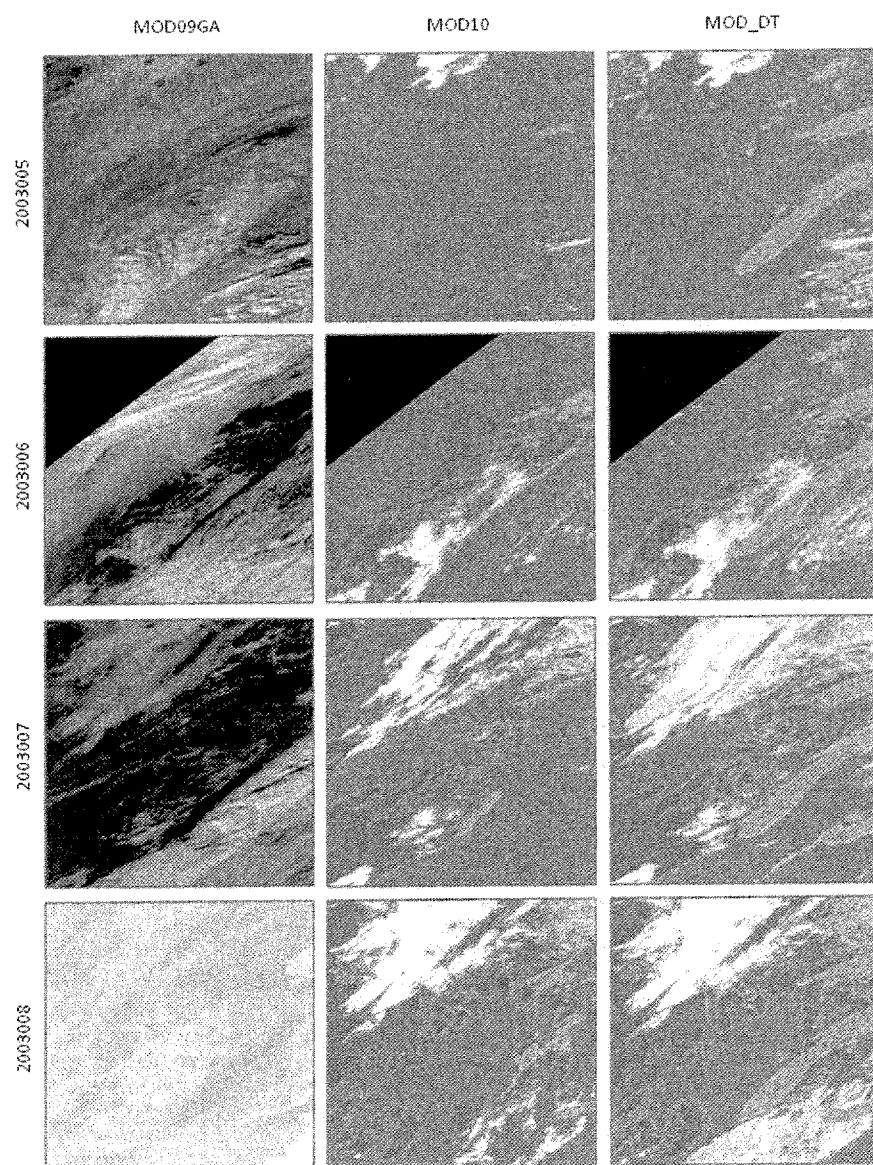
Figure 3. MOD10 and MOD_DT snow maps for the time series 1-10 January 2003. False color images of MOD09GA (first column) made from MODIS bands 1, 4, 6 in RGB image layers, respectively. Refer to Fig. 2 for map orientation.
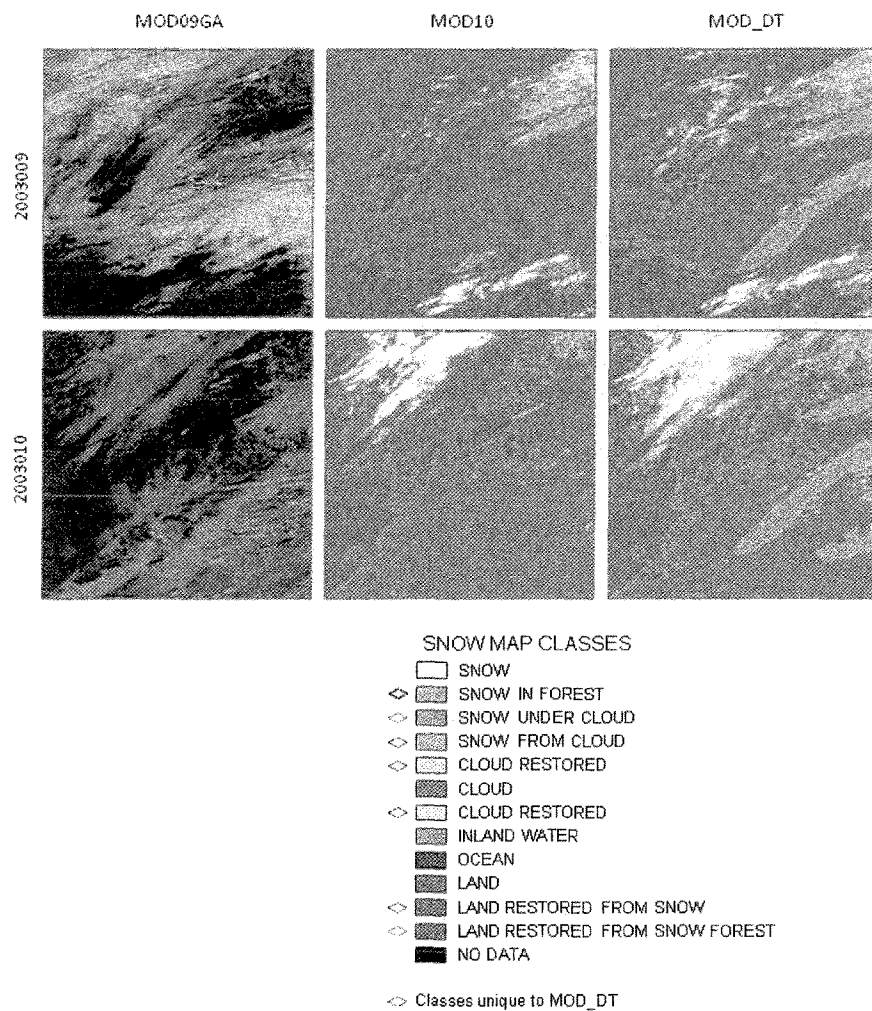
Figure 3. MOD10 and MOD_DT snow maps for the time series 1-10 January 2003. False color images of MOD09GA (first column) made from MODIS bands 1, 4, 6 in RGB image layers, respectively. Refer to Fig. 2 for map orientation.

Figure 3. MOD10 and MOD_DT snow maps for the time series 1-10 January 2003. False color images of MOD09GA (first column) made from MODIS bands 1, 4, 6 in RGB image layers, respectively. Refer to Fig. 2 for map orientation.
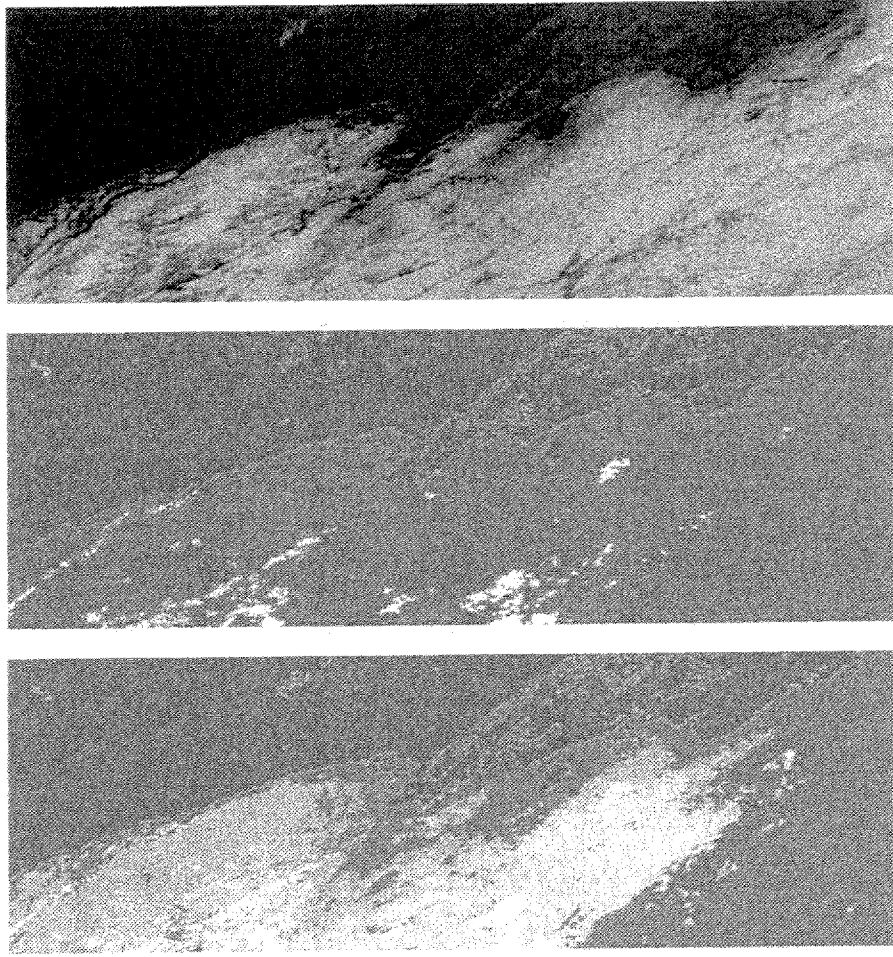
Figure 4. Example of reduced snow/cloud confusion in a MOD_DT snow map. Images are from 3 January 2003, lower left corner of tile h14v03 (~562 x 152 km area); top MOD09GA, middle MOD10, and bottom MOD_DT. In MOD_DT much of the area visually interpreted as snow covered ground on the left side (MOD09GA) was erroneously flagged as snow in MOD10, but much of that cloud error was changed to snow cover (pale blue) in MOD_DT, bottom image. Colors are the same as in Figure 3.

CONCLUSION

A slightly more accurate snow map was made from the decision tree algorithm primarily by decreasing snow/cloud confusion using a branch of nodes to alleviate snow/cloud confusion. Nodes were used to discern snow from cloud and to affirm a result or to restore a previous result in order to alleviate snow/cloud confusion over snow-covered surfaces while minimizing concomitant errors of snow/cloud confusion in other situations. The classified observation from

the terminal node of a branch indicates under what conditions a decision was reached. There are more classes of snow cover in the MOD_DT snow map compared to the MOD10 snow map that can be utilized by a user. Though the improvement in snow mapping from improved snow/cloud discrimination was slight, 0 – 5% of snow covered land in a scene; it does enhance use of the snow maps for monitoring snow, for example for building snow cover depletion curves.

The capability to improve snow mapping with the MOD_DT will continue to be investigated with the objective of further decreasing snow cloud confusion to improve our ability to detect snow cover under non-optimal conditions. A possibility being explored is to use the MOD09GA internal cloud flag in combination with the MOD35 cloud flag to improve the accuracy of cloud masking and development of nodes for snow classification. Initial results using the MO09GA and MOD35 cloud flags have been encouraging in regard to masking of clouds over snow. The MOD_DT algorithm will be evaluated using many more scenes from different locations and dates, before a stable algorithm will be finalized.

REFERENCES

Ackerman S, Frey R, Strabala K, Liu Y, Gumley L, Baum B and Menzel P. 2010. *Discriminating clear-sky from cloud with MODIS: Algorithm Theoretical Basis Document (MOD35), Version 6*, ATBD Reference Number: ATBD-MOD-06, modis-atmos.gsfc.nasa.gov/reference_atbd.html (11 August 2010).

Friedl MA and Brodley CE. 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of the Environment* **61**:399-409.

Hall DK and Riggs GA. 2007. Accuracy assessment of the MODIS snow-cover products. *Hydrological Processes* **21**:1534–1547.

Hansen M, Dubayah R and DeFries R. 1996. Classification trees: an alternative to traditional land cover classifiers. *International Journal Remote Sensing* **17**(5):1075-1081

Pouliot D, Latifovic R, Fernandes R and Olthof I. 2009. Evaluation of annual forest disturbance monitoring using a static decision tree approach and 250 m MODIS data. *Remote Sensing of Environment* **113**:1749-1759.

Riggs GA and Hall DK. 2002. Reduction of cloud obscuration in the EOS MODIS Snow Data Product. *59th Eastern Snow Conference, 5 - 7 June 2002, Stowe, VT.* Published online in 60th Eastern Snow Conference Proceedings as an addendum. www.easternsnow.org/proceedings/2003/proceedings_index.html.

Riggs GA, Hall DK and Salomonson,V.V. 2006. *MODIS snow products user guide.* http://modis-snow-ice.gsfc.nasa.gov/sug_c5.pdf (11 August 2010).

Simard M, Saatchi SS, DeGrandi G. 2000. The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical forest, *IEEE Transactions on Geoscience and Remote Sensing* **38**(5):2310-2321.